

**PENGEMBANGAN TES PILIHAN GANDA MENGGUNAKAN
MODEL RASCH MATERI GERAK LURUS
KELAS X PONTIANAK**

ARTIKEL PENELITIAN



**OLEH:
MUHAMMAD RAMADHAN RIDHO NOVINDA
NIM. F1051141055**

**PROGRAM STUDI PENDIDIKAN FISIKA
JURUSAN PENDIDIKAN MATEMATIKA DAN IPA
FAKULTAS KEGURUAN DAN ILMU PENDIDIKAN
PONTIANAK
2019**

LEMBAR PERSETUJUAN

PENGEMBANGAN TES PILIHAN GANDA MENGGUNAKAN MODEL RASCH MATERI GERAK LURUS KELAS X PONTIANAK

ARTIKEL PENELITIAN

MUHAMMAD RAMADHAN RIDHO NOVINDA
NIM. F1051141055

Pembimbing I

Disetujui,

Pembimbing II



Dr. Haratua Tiur Maria S, M.Pd
NIP. 196702221991012001



Hamdani, M.Pd
NIP. 198506052008121001

Mengetahui,

Dekan FKIP

Ketua Jurusan P.MIPA



Dr. H. Martono, M.Pd
NIP. 196803161994031014



Dr. Ahmad Yani. T
NIP. 196604011991021001

PENGEMBANGAN TES PILIHAN GANDA MENGGUNAKAN MODEL RASCH MATERI GERAK LURUS KELAS X PONTIANAK

Muhammad Ramadhan Ridho Novinda, Haratua Tiur Maria Silitonga, Hamdani

Program Studi Fisika FKIP Untan Pontianak

Email: ramadhan183@ymail.com

Abstract

This study aimed to create multiple choice objective test on linear motion to measure the learning outcomes of X grade high school students in Pontianak City. The method used was Research and Development (R & D) with the Four-D models. Development tests using Mardapi steps which has 9 steps. Consisting 40 item of tests. Trial questions involved 33 students in small-scale trials, and 204 students in large-scale trials. Item characteristics were analyzed using the Winstep program with the Rasch Model. Based on the analysis of the data, it can be concluded that the test developed worthy of use to measure student learning outcomes are: 1) items made with simple language with a readability level of 6.60, 2) a high level of content validity which is equal to 0.75, 3) all items spread evenly on person-item maps, 4) there are 28 items which are included in the easy to hard category, 5) the value of unidimensionality which includes enough categories is 31%, 6) the value of items reliability that belong to the special category is equal to 0,98. Based on the result of the analysis, there are 26 items that are worthy of use.

Keywords: *Linear Motion, Rasch Model, Test Development.*

PENDAHULUAN

Teori pengukuran klasik dikenal juga sebagai teori ujian klasik (*classical test theory*), memperkenalkan tiga konsep: skor tes (sering disebut skor yang diamati), skor kebenaran, dan skor kesalahan. Dalam kerangka teoritis itu, berbagai bentuk model telah dirumuskan. Sebagai contoh, dalam apa yang sering disebut sebagai "*classical test model*", model linear sederhana yang dipostulasikan mengaitkan yang dapat diamati skor tes (X) untuk jumlah dua tidak teramati (atau sering disebut laten) variabel, skor benar (T) dan skor kesalahan (E), itu adalah, $X = T + E$ (Hambleton and Jones, 1993).

Pada teori klasik, taraf sukar butir bergantung (*dependent*) kepada kemampuan responden. Bagi responden berkemampuan tinggi, butir menjadi tidak sukar (mudah). Bagi responden berkemampuan rendah, butir menjadi sukar. Pada butir tidak sukar (mudah), tampak kemampuan responden menjadi tinggi. Pada butir sukar, tampak

kemampuan responden menjadi rendah. Taraf sukar butir bergantung pada kemampuan responden. Butir yang sama akan terasa berat bagi mereka yang berkemampuan rendah dan terasa ringan bagi mereka yang berkemampuan tinggi. Teori pengukuran klasik (teori ujian klasik) tidak dapat digunakan untuk pencocokan kemampuan responden dengan taraf sukar butir (Sudaryono, 2012).

Qasem (2013) menjelaskan salah satu kelemahan teori pengukuran klasik yaitu teori ini mengasumsikan bahwa nilai tes yang mempresentasikan sifat atau kemampuan harus dalam fungsi linear stabil, jika skor individu mengalami peningkatan dalam tes jumlah kemampuannya harus meningkat juga. Namun, beberapa individu dengan kemampuan tinggi terkadang mendapat skor rendah pada tes, dan mungkin yang sebaliknya akan terjadi bagi mereka yang memiliki kemampuan rendah.

Sumintono dan Widhiarso (2015) juga memaparkan beberapa kelemahan teori

pengukuran klasik yang hanya berpatokan pada skor mentah (*raw score*), yaitu: (a) skor mentah pada dasarnya bukanlah hasil pengukuran. Lebih tepatnya skor mentah adalah jumlah jawaban benar dari soal yang dikerjakan peserta didik, (b) skor mentah adalah informasi awal. Skor mentah juga biasanya dinyatakan dalam persentase (%) yang tidak lain hanyalah ringkasan data berupa angka, tetapi tidak memberikan data suatu pengukuran, (c) skor mentah memiliki makna kuantitatif yang lemah. Makna kuantitatif dari skor mentah yang didapatkan akan berbeda, bergantung pada banyaknya soal, sedangkan persentase jawaban betul selalu bergantung pada tingkat kesulitan soal, (d) skor mentah tidak menunjukkan kemampuan seseorang terhadap tugas tertentu, skor mentah juga tidak bisa banyak menjelaskan tingkat kesulitan soalnya, (e) skor mentah dan persentase jawaban benar tidak selalu bersifat linear, dalam sebuah tes yang bersifat linear, peserta didik yang memiliki skor 15 (skala 0 hingga 100) selalu memiliki kemampuan lebih tinggi dibanding yang memiliki skor 10. Namun, secara empirik terkadang keduanya memungkinkan memiliki kemampuan yang sama.

Dari beberapa penjelasan tentang kelemahan teori pengukuran klasik di atas, dapat disimpulkan bahwa teori pengukuran klasik kurang tepat jika digunakan untuk mengukur kemampuan peserta didik serta menganalisis butir soal. Oleh karena itu, dibutuhkan sebuah teori pengukuran baru untuk mengatasi kelemahan teori pengukuran klasik. Penelitian ini menyarankan penggunaan teori pengukuran modern atau biasa disebut *Item Response Theory (IRT)* untuk menganalisis butir soal sehingga kemampuan peserta didik dapat terukur dengan baik.

Teori pengukuran modern dikenal sebagai teori ujian modern (*modern test theory*). Teori ini berfokus pada butir, yang bertentangan dengan teori klasik. *IRT* memodelkan respon setiap peserta untuk setiap item dalam ujian. Item yang mencakup semua jenis butir. Pertanyaan pilihan ganda

yang mempunyai jawaban salah dan benar (Qasem, 2013).

IRT adalah teori statistik umum tentang memeriksa item dan menguji kinerja serta bagaimana kinerja berhubungan dengan kemampuan yang diukur oleh item dalam tes. Item respon dapat bersifat diskrit atau kontinu dan dapat berupa skor dikotomi atau politomi, kategori skor item dapat diurutkan atau tidak diurutkan, bisa ada satu kemampuan atau banyak kemampuan yang mendasari kinerja tes, dan ada banyak cara (yaitu model) di mana hubungan antara item respons dan kemampuan dasar atau kemampuan dapat ditentukan. Dalam kerangka kerja *IRT* umum, banyak model yang sudah diformulasikan dan diaplikasikan pada data uji nyata namun yang paling terkenal adalah Model Rasch (Qasem, 2013).

Pemodelan Rasch bertujuan mengembangkan pengukuran yang objektif. Dalam konteks pemodelan Rasch, antonim dari “pengukuran yang objektif” (*objective measurement*) bukanlah “pengukuran subjektif”, melainkan pengukuran yang hasilnya bergantung pada siapa yang diukur (*test-dependent scoring*). Persentase atau jumlah jawaban benar pada sebuah tes matematika bergantung pada subjek yang diukur (*sample dependent*) yang bersifat deskriptif dan berlaku untuk semua subjek tersebut (Sumintono dan Widhiarso, 2013).

Untuk memastikan pengukuran menjadi objektif maka model pengukuran haruslah memenuhi lima syarat ini: (a) memberikan ukuran yang linear, (b) mengatasi data yang hilang, (c) melakukan proses estimasi yang tepat, (d) menemukan yang tidak tepat (*misfits*) atau tidak umum (*outliers*), (e) memberikan instrumen pengukuran yang independen dari parameter yang diteliti (Sumintono dan Widhiarso, 2013).

Untuk saat ini hanya pemodelan pengukuran Rasch-lah yang memenuhi kelima syarat tersebut. Pengukuran objektif menghasilkan data yang terbebas dari jenis subjek, karakteristik penilai (*rater*) dan karakteristik alat ukur (Sumintono dan Widhiarso, 2013). Inilah yang menjadi acuan

penelitian ini menggunakan pemodelan Rasch untuk menganalisis instrumen tes.

Ada banyak jenis tes yang dapat dirancang oleh guru untuk mengukur kemampuan peserta didik diantaranya adalah tes pilihan berganda (*multiple choice*), tes uraian (*essay*), tes benar-salah (*true-flase test*), dan tes menjodohkan (*matching test*) (Arikunto, 2012).

Berdasarkan hasil wawancara dengan guru di SMA Pontianak, kadang kala instrumen tes yang dirancang tidak melalui proses validasi, reliabilitas serta tidak didahului uji coba sehingga instrumen tes ini dapat dikatakan tidak valid. Instrumen tes yang dirancang berbentuk pilihan ganda yang akan digunakan untuk ulangan harian. Namun, karena instrumen tes yang dirancang tidak valid dapat menyebabkan kemampuan peserta didik yang terukur juga menjadi tidak valid.

Dari uraian latar belakang di atas, penelitian ini dilakukan untuk mengembangkan instrumen tes pilihan berganda yang sesuai dengan Model Rasch pada materi Gerak Lurus. Instrumen tes yang dikembangkan tentu saja akan melalui proses uji coba, serta menghitung validitas dan reliabilitas sehingga dapat memenuhi karakteristik-karakteristik tes yang baik berdasarkan Model Rasch.

Penelitian ini bertujuan untuk menghasilkan instrumen tes pilihan ganda untuk mengukur kemampuan peserta didik pada materi gerak lurus kelas X yang memenuhi karakteristik-karakteristik Model Rasch.

METODE PENELITIAN

Peneliti menggunakan langkah-langkah 4-D untuk melakukan penelitian dan pengembangan menurut Mardapi (2012). Langkah-langkah 4-D meliputi:

Tahap pertama yaitu tahap pendefinisian (*Define*), menetapkan dan mendefinisikan syarat-syarat pembelajaran diawali dengan analisis tujuan dari batasan materi yang dikembangkan. Terdiri dari, (1) Analisis Ujung Depan, (2) Analisis Peserta Didik, (3)

Analisis Konsep, (4) Analisis Tugas, (5) Penentuan Tujuan Pembelajaran.

Tahap kedua yaitu tahap perancangan (*Design*), menyiapkan instrumen perangkat pembelajaran. Tahap perencanaan terdiri dari empat langkah yaitu, (1) Menyusun Kisi-Kisi, (2) Menentukan Bentuk Dan Jumlah Soal Yang Sesuai, (3) Menulis Soal, (4) Menentukan Panjang Tes.

Tahap ketiga yaitu tahap pengembangan (*Develope*), menghasilkan perangkat pembelajaran yang sudah direvisi berdasarkan masukan dari pakar. Tahap pengembangan meliputi, (1) Menelaah Tes, (2) Melakukan Uji Coba Tes, (3) Menganalisis Tes, (4) Memperbaiki Tes, (5) Merakit Tes.

Tahap keempat yaitu tahap penyebaran (*Desseminate*), penggunaan perangkat yang telah dikembangkan pada skala yang lebih luas serta menguji efektivitas penggunaan perangkat di dalam KBM. Tahap penyebaran meliputi (1) Melaksanakan Tes dan (2) Menafsirkan Hasil Tes.

Populasi dalam penelitian ini adalah peserta didik kelas X SMA di Kota Pontianak. Sedangkan sampel pada penelitian ini diambil berdasarkan ukuran sampel dalam pemodelan Rasch dengan nilai $\pm 0,5$ logit serta tingkat kecepecaayaan 99% yaitu kisaran sampel berjumlah 108-243 sampel. Sampel diambil dari rerata nilai Ujian Nasional (UN) IPA tinggi, sedang, dan rendah pada tahun 2017 di Kota Pontianak.

Agar data yang diperoleh memenuhi karakteristik Model Rasch, data tersebut dianalisis menggunakan *software* Winstep. Dalam penelitian ini, analisis data berupa unidimensionalitas, reliabilitas, peta person-item, dan tingkat kesukaran butir soal.

HASIL DAN PEMBAHASAN

Hasil

Penelitian ini dilaksanakan di Sekolah Menengah Atas (SMA) Kota Pontianak yang berlangsung di kelas X SMA Negeri 3 Pontianak, kelas X SMA Negeri 5 Pontianak, kelas X SMA Negeri 7 Pontianak, dan kelas X SMA Negeri 8 Pontianak pada tahun ajaran 2018/2019. Penelitian ini berlangsung dua tahap, yaitu uji skala kecil dan uji skala

besar. Uji skala kecil dilakukan di kelas X MIA 4 SMA Negeri 8 Pontianak dengan jumlah peserta didik 33 orang. Uji skala besar dilakukan di kelas X MIA 5 dan X MIA 6 SMA Negeri 3 Pontianak, kelas X MIA 1 dan X MIA 3 SMA Negeri 5 Pontianak, serta X MIA 1 dan X MIA 4 SMA Negeri 7 Pontianak dengan jumlah peserta didik 204 orang.

Tahap pertama dalam pengembangan tes adalah *Define* (Tahap Pendefinisian). Langkah dalam tahap ini yaitu: (1) Analisis ujung depan. Masalah yang diangkat dalam penelitian ini adalah kelayakan instrumen tes yang digunakan untuk mengukur kemampuan peserta didik. (2) Analisis peserta didik. Peneliti menggunakan rerata nilai UN mata pelajaran Fisika untuk menentukan kemampuan akademik peserta didik SMA Negeri kota Pontianak. (3) Analisis konsep. Konsep yang digunakan untuk mengembangkan instrumen tes berupa materi gerak lurus dengan sub-materi GLB, GLBB, gerak vertikal, gerak jatuh bebas. (4) Analisis tugas. Analisis yang dilakukan pada tahap ini adalah analisis Kompetensi Inti (KI) dan Kompetensi Dasar (KD) terkait instrumen tes yang akan dikembangkan. (5) Penentuan tujuan pembelajaran, Menentukan indikator instrumen tes yang harus dicapai oleh peserta didik untuk setiap butir soal berdasarkan KI, KD serta materi yang telah ditentukan.

Tahap kedua yaitu *Design* (Tahap Perancangan). Langkah dalam tahap ini yaitu: (1) Penyusunan kisi-kisi dan indikator soal pada materi gerak lurus. (2) Menentukan bentuk dan jumlah soal yang sesuai. Soal yang dirancang berupa soal pilihan ganda dengan total 40 butir dengan lima alternatif jawaban untuk setiap butir soal. (3) Menulis soal. Tes dirancang dengan tingkat taksonomi bloom C2-C5 yang telah direvisi menurut Anderson dan Krathwohl. (4) Menentukan panjang tes. Menentukan waktu yang disediakan berdasarkan tingkat perkembangan peserta didik dan jenjang pendidikan.

Tahap ketiga yaitu *Develop* (Tahap Pengembangan). Langkah dalam tahap ini yaitu: (1) Menelaah tes. Instrumen tes yang

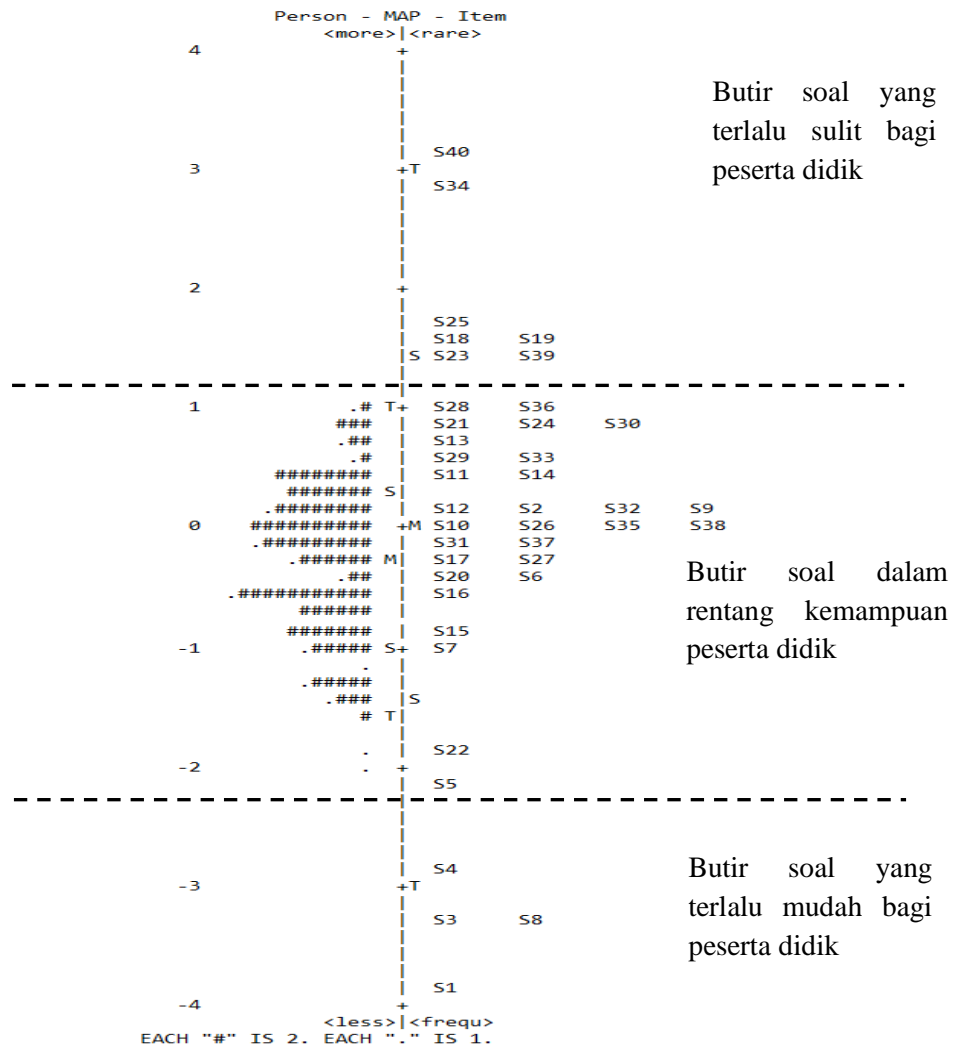
telah dirancang kemudian dihitung tingkat keterbacaan menggunakan microsoft excel. Hasil perhitungan tingkat keterbacaan yang telah dirancang harus lebih dari sama dengan 6. Rata-rata tingkat keterbacaan instrumen tes adalah 6,60. Instrumen tes kemudian divalidasi isi oleh tujuh orang validator yang terdiri dari dua orang dosen fisika dan lima orang guru fisika SMA. Validasi bertujuan agar tes yang dirancang sesuai dengan teori dan materi gerak lurus di sekolah. Hasil validasi isi dihitung menggunakan rumus Aiken dan didapat rata-rata validasi isi oleh validator sebesar 0,75 yang menunjukkan instrumen tes yang dikembangkan berada pada kategori tinggi sehingga layak untuk digunakan.

(2) Melakukan uji coba tes atau uji coba skala kecil. Dilakukan di kelas X MIA 4 SMA Negeri 8 Pontianak, uji coba skala kecil bertujuan untuk mengetahui waktu pengerjaan, alternatif jawaban, dan penggunaan bahasa yang nantinya akan diperbaiki untuk melakukan uji skala besar. (3) Menganalisis butir soal. Dari segi waktu, peserta didik dapat menyelesaikan instrumen tes dalam waktu lebih kurang 70 menit sehingga 2 jam pelajaran (90 menit) sudah cukup untuk melakukan uji skala besar. Pada alternatif jawaban tidak terdapat masalah pada peserta didik sedangkan pada bahasa yang digunakan sebagai pertanyaan soal membingungkan beberapa peserta didik sehingga perlu diperbaiki. (4) Memperbaiki tes. Perbaikan tes dilakukan dengan menganalisis butir soal yang perlu diperbaiki berdasarkan hasil validitas isi oleh validator serta hasil analisis pada uji coba skala kecil. (5) Merakit tes. Setelah menganalisis dan memperbaiki kesalahan pada butir soal, Instrumen tes dirakit kembali sehingga menghasilkan instrumen tes pilihan ganda pada materi Gerak Lurus yang sudah digunakan pada skala besar dengan tetap mempertahankan jumlah butir soal.

Tahap keempat yaitu *Desseminate* (Tahap Penyebaran). Langkah dalam tahap ini yaitu: (1) Melaksanakan tes atau uji skala besar. Tes dilaksanakan di kelas X MIA 5 dan 6 SMA Negeri 3 Pontianak, kelas X MIA

1 dan 3 SMA Negeri 5 Pontianak, serta kelas X MIA 1 dan 4 SMA Negeri 7 Pontianak. (2) Menafsirkan hasil tes. Hasil uji skala besar dianalisis dengan *software* Winstep

kemudian diseleksi butir soal yang layak digunakan untuk mengukur kemampuan peserta didik.



Gambar 1. Person-Item Map

Tabel 1. Tingkat Kesukaran Butir Soal

No. Soal	Tingkat Kesukaran	INFIT		OUTFIT	
		MNSQ	ZSTD	MNSQ	ZSTD
1	-3.82	.96	.0	.55	-1.1
2	.14	1.00	.1	1.00	.0
3	-3.24	.98	.0	.82	-.5
4	-2.79	.98	-.1	.94	-.2
5	-2.10	.90	-.7	.74	-1.7
6	-.47	.93	-1.6	.94	-1.2
7	-1.07	1.10	1.6	1.14	1.8

No. Soal	Tingkat Kesukaran	INFIT		OUTFIT	
		MNSQ	ZSTD	MNSQ	ZSTD
8	-3.34	.95	-.1	.84	-.4
9	.13	.88	-2.5	.86	-2.4
10	-.06	.97	-.6	.96	-.8
11	.49	1.16	2.5	1.22	2.5
12	.19	.81	-4.2	.77	-4.1
13	.68	1.05	.8	1.09	1.0
14	.50	1.06	.9	1.05	.6
15	-.80	1.19	3.5	1.27	4.2
16	-.52	.95	-1.3	.94	-1.2
17	-.35	1.29	6.5	1.33	6.5
18	1.51	1.03	.3	1.21	1.2
19	1.56	.96	-.2	.92	-.4
20	-.40	.80	-5.0	.78	-5.0
21	.83	1.15	1.9	1.27	2.4
22	-1.87	.96	-.4	.87	-.9
23	1.40	.92	-.6	.89	-.7
24	.81	1.05	.6	1.03	.3
25	1.67	1.02	.2	1.09	.5
26	.07	1.13	2.6	1.13	2.2
27	-.33	1.09	2.1	1.09	1.9
28	.98	1.08	.8	1.12	1.0
29	.57	.91	-1.3	.94	-.7
30	.82	1.02	.3	1.06	.6
31	-.19	.77	-6.0	.75	-5.8
32	.19	.80	-4.3	.76	-4.2
33	.56	.84	-2.5	.78	-2.7
34	2.83	1.04	.3	1.30	.9
35	.00	1.17	3.4	1.25	4.0
36	.98	.99	.0	.95	-.3
37	-.13	.91	-2.2	.91	-2.0
38	-.04	1.02	.5	1.03	.6
39	1.47	1.06	.5	1.23	1.3
40	3.15	1.04	.2	2.05	2.0

Tabel 2. Hasil Analisis Butir Soal

No. Soal	Kriteria						Keterangan
	1	2	3	4	5	6	
1.	6,79	0,76	31%	0,98	-	-	DIBUANG
2.	6,26	0,76	31%	0,98	✓	✓	DIPAKAI
3.	6,18	0,75	31%	0,98	-	-	DIBUANG
4.	6,30	0,75	31%	0,98	-	-	DIBUANG
5.	7,14	0,77	31%	0,98	-	-	DIBUANG
6.	6,05	0,77	31%	0,98	✓	✓	DIPAKAI
7.	6,64	0,75	31%	0,98	✓	-	DIBUANG
8.	6,12	0,75	31%	0,98	✓	-	DIBUANG
9.	6,56	0,75	31%	0,98	✓	✓	DIPAKAI

No. Soal	Kriteria						Keterangan
	1	2	3	4	5	6	
10.	6,11	0,75	31%	0,98	✓	✓	DIPAKAI
11.	6,16	0,68	31%	0,98	✓	✓	DIPAKAI
12.	6,87	0,75	31%	0,98	✓	✓	DIPAKAI
13.	6,12	0,75	31%	0,98	✓	✓	DIPAKAI
14.	6,35	0,75	31%	0,98	✓	✓	DIPAKAI
15.	6,51	0,75	31%	0,98	✓	✓	DIPAKAI
16.	6,64	0,75	31%	0,98	✓	✓	DIPAKAI
17.	6,62	0,76	31%	0,98	✓	✓	DIPAKAI
18.	6,40	0,76	31%	0,98	-	-	DIBUANG
19.	6,52	0,75	31%	0,98	-	-	DIBUANG
20.	7,26	0,75	31%	0,98	✓	✓	DIPAKAI
21.	6,27	0,76	31%	0,98	✓	✓	DIPAKAI
22.	6,96	0,76	31%	0,98	✓	-	DIBUANG
23.	8,31	0,77	31%	0,98	-	-	DIBUANG
24.	7,11	0,75	31%	0,98	✓	✓	DIPAKAI
25.	7,40	0,76	31%	0,98	-	-	DIBUANG
26.	6,71	0,76	31%	0,98	✓	✓	DIPAKAI
27.	6,67	0,76	31%	0,98	✓	✓	DIPAKAI
28.	6,25	0,74	31%	0,98	✓	✓	DIPAKAI
29.	6,64	0,76	31%	0,98	✓	✓	DIPAKAI
30.	7,50	0,76	31%	0,98	✓	✓	DIPAKAI
31.	6,27	0,75	31%	0,98	✓	✓	DIPAKAI
32.	6,03	0,76	31%	0,98	✓	✓	DIPAKAI
33.	7,16	0,76	31%	0,98	✓	✓	DIPAKAI
34.	6,16	0,75	31%	0,98	-	-	DIBUANG
35.	6,22	0,76	31%	0,98	✓	✓	DIPAKAI
36.	6,48	0,75	31%	0,98	✓	✓	DIPAKAI
37.	6,55	0,76	31%	0,98	✓	✓	DIPAKAI
38.	7,01	0,76	31%	0,98	✓	✓	DIPAKAI
39.	6,42	0,76	31%	0,98	-	-	DIBUANG
40.	6,30	0,75	31%	0,98	-	-	DIBUANG

Keterangan Kriteria:

1. Tingkat keterbacaan (Minimal 6,0)
2. Validitas isi oleh validator (Minimal 0,61)
3. Unidimensionalitas (Minimal 20%)
4. Reliabilitas item (Minimal 0,67)
5. Peta person-item (✓ = Memenuhi, - = Tidak memenuhi)
6. Tingkat kesukaran soal (✓ = Memenuhi, - = Tidak memenuhi)

Pembahasan

Unidimensionalitas adalah ukuran yang penting untuk mengevaluasi apakah instrumen tes pilihan ganda yang dikembangkan mampu mengukur apa yang seharusnya diukur. Analisis Model Rasch menggunakan analisis komponen utama (*Partial Component Analysis*) dari residual, yaitu mengukur sejauh mana keragaman dari

instrumen tes yang dikembangkan mengukur apa yang seharusnya diukur. Dari hasil analisis ini diperoleh hasil pengukuran *Raw variance* sebesar 31%. Hal ini menunjukkan bahwa persyaratan unidimensionalitas minimal sebesar 20% dapat dipenuhi. Dengan demikian instrumen tes pilihan ganda yang dikembangkan dalam penelitian ini

cukup valid untuk mengukur kemampuan peserta didik.

Reliabilitas adalah sejauh mana hasil suatu pengukuran dapat dipercaya (Azwar, 2013). Reliabilitas instrumen tes dalam penelitian ini dilihat dari *real item reliability* karena nilainya lebih konservatif dibandingkan *model item reliability* (Boone, dkk, 2014). Dari hasil analisis program Winstep diperoleh nilai *real item reliability* sebesar 0,98 yang termasuk ke dalam kategori istimewa. Melalui *item reliability* yang tinggi ini, dapat disimpulkan bahwa instrumen tes yang dikembangkan sudah terdapat beberapa item yang lebih sulit dan beberapa item yang lebih mudah serta konsistensi dari kesimpulan ini dapat diharapkan (Bond dan Fox, 2015).

Peta person-item menunjukkan sebaran kemampuan peserta didik (sebelah kiri) dan tingkat kesukaran butir soal (sebelah kanan) pada interval yang sama. Melalui peta ini, peneliti mengambil butir soal yang dapat dijangkau oleh kemampuan peserta didik. Pada peta, peneliti mengambil 3 analisis data yaitu butir soal dengan tingkat kesukaran melebihi tingkat kemampuan peserta didik tertinggi, butir soal dengan tingkat kesukaran kurang dari kemampuan peserta didik terendah, dan butir soal dengan tingkat kesukaran yang dapat dijangkau oleh kemampuan peserta didik tertinggi maupun terendah.

Data pertama hasil analisis adalah butir soal dengan tingkat kesukaran melebihi tingkat kemampuan peserta didik tertinggi. Terdapat tujuh butir soal dengan tingkat kesukaran melebihi tingkat kemampuan peserta didik tertinggi, butir soal yang terlalu sulit dapat membuat beberapa peserta didik frustrasi untuk menyelesaikan butir tersebut serta mengurangi motivasi peserta didik dalam belajar. Karena tujuan pengembangan instrumen tes dalam penelitian ini adalah tes formatif, ketujuh butir ini tidak dapat digunakan. Namun, butir soal ini masih dapat digunakan jika tujuan tes bersifat seleksi. Butir soal tersebut adalah butir nomor 40, 34, 25, 18, 19, 23, dan 39.

Data kedua hasil analisis adalah butir soal dengan tingkat kesukaran yang lebih rendah dari kemampuan peserta didik terendah. Terdapat lima butir soal pada tingkat kesukaran ini, butir soal yang terlalu mudah tidak akan merangsang kemampuan berpikir peserta didik yang dapat mengakibatkan kemampuan berpikir peserta didik tidak berkembang. Dapat disimpulkan, kelima butir soal ini lebih baik tidak digunakan dalam tes formatif. Butir soal tersebut adalah butir nomor 5, 4, 3, 8, dan 1.

Data terakhir hasil analisis dari peta variabel adalah butir soal dengan tingkat kesukaran yang dapat dijangkau oleh kemampuan peserta didik tertinggi maupun terendah. Terdapat 28 butir soal yang didapat dari data ini. Butir-butir ini dapat dipertahankan karena 28 butir soal ini masih dapat dikerjakan oleh seluruh peserta didik, baik yang memiliki kemampuan akademik rendah maupun tingkat kemampuan akademik yang tinggi. Butir soal tersebut adalah butir nomor 28, 36, 21, 24, 30, 13, 29, 33, 11, 14, 12, 2, 32, 9, 10, 31, 37, 17, 27, 20, 6, 16, 15, 7, dan 22.

Dari analisis di atas, dapat disimpulkan bahwa butir soal yang dikembangkan sudah baik karena dapat menjangkau seluruh abilitas peserta didik. Namun, terdapat beberapa butir soal yang terlalu sukar dan terlalu mudah. Sehingga perlu dilakukan analisa lanjutan untuk memilih butir soal yang layak digunakan untuk mengukur tingkat abilitas peserta didik.

Keunggulan lain dari Model Rasch dibanding metode lainnya, khususnya dari teori tes klasik, kemampuan melakukan prediksi terhadap data yang hilang (*missing data*), yang didasarkan kepada pola respon yang sistematis. Dalam model lain biasanya mengestimasi data yang hilang dengan nilai nol (0), sedangkan Rasch model akan menghasilkan prediksi mana kemungkinan nilai terbaik dari data yang hilang tersebut. Dengan demikian data yang diperoleh seolah-olah sebagai data yang lengkap dan lebih akurat dalam analisis statistik selanjutnya (Akhtar, 2017).

Dalam penelitian ini, jawaban kosong yang tidak dikerjakan oleh peserta didik pada beberapa butir soal tidak dianggap sebagai jawaban yang salah, tetapi dianggap sebagai data yang hilang atau *missing data*. Hal ini dilakukan untuk memanfaatkan kelebihan Model Rasch yang tidak ada pada teori tes klasik yaitu dapat memprediksi *missing data*.

Untuk memastikan butir soal yang digunakan *fit* atau sesuai dengan Model Rasch beberapa poin yang dapat dianalisis adalah *measure* (menunjukkan tingkat kesulitan item), *infit-outfit MNSQ* dan *ZSTD* (menunjukkan apakah item kita sesuai dengan Model Rasch) (Akhtar, 2017).

Measure, butir soal yang akan diambil pada penelitian ini adalah butir soal dengan interval nilai $-1 < measure < 1$ artinya yang termasuk kategori mudah dan sulit (Sumintono dan Widhiarso, 2013).

MNSQ, Mean-Square Fit Statistic (MNSQ) memperlihatkan ukuran ke-acak-an, yaitu jumlah distorsi dalam sistem pengukuran. Nilai yang diharapkan adalah antara 0,5 s.d 1,5; jika nilai kurang dari nilai tersebut, mengindikasikan hal itu terlalu mudah ditebak; sedangkan nilai yang lebih besar mengindikasikan tidak mudah diprediksi (Sumintono dan Widhiarso, 2013).

ZSTD, Standarized Fit Statistic (ZSTD) adalah uji-t untuk hipotesis, ‘apakah data sesuai (fit) dengan model?’ Hasilnya adalah nilai-z yaitu penyimpangan unit. Ini menjelaskan ketidaksesuaian dari data, yaitu signifikansinya jika data memang sesuai dengan model. Nilai yang diharapkan adalah antara -1,9 s.d 1,9 (Sumintono dan Widhiarso, 2013).

Ketika akan mengevaluasi item, Boone. dkk (2014) menyarankan untuk mengidentifikasi *outfit* dan lebih khusus lagi

item *outfit* MNSQ karena statistik outfit lebih sensitif terhadap *outlier* dan memiliki perhitungan yang lebih lazim. Sensitifitas statistik *outfit* juga membuat item lebih mudah untuk diidentifikasi dan memperbaiki masalah kecocokan. Terlebih lagi, Linacre (2012) dalam Boone. dkk (2014) menyatakan secara khusus untuk tujuan pelaporan bahwa hanya outfit yang perlu dilaporkan, “kecuali data terkontamiasi dengan *outlier* yang tidak relevan”. Maka pelaporan *infit* mungkin diperlukan.

Oleh karena penelitian ini terdapat banyak peserta didik yang tidak mengisi lembar jawaban, maka nilai *infit* MNSQ juga perlu dianalisis dan dilaporkan. Salah satu statistik fit item yang utama adalah *infit mean square* (INFIT MNSQ). *Infit mean square* mengukur konsistensi kecocokan peserta didik dengan kurva karakteristik item untuk setiap item dengan pertimbangan yang diberikan kepada person yang dekat dengan tingkat probabilitas 0,5 (Alagumalai. dkk, 2005).

Pertama, periksa nilai MNSQ untuk mengevaluasi kecocokan, selama nilai MNSQ berada dalam rentang kesesuaian yang diterima, nilai ZSTD diabaikan (Boone. dkk, 2014). Karena nilai *infit* MNSQ dan *outfit* MNSQ pada penelitian ini berada pada rentang yang dapat diterima (memiliki nilai antara 0.5 s.d 1,5), maka nilai ZSTD diabaikan.

Berdasarkan kriteria di atas, butir soal yang memenuhi seluruh kriteria dan layak digunakan untuk mengukur kemampuan peserta didik pada materi Gerak Lurus berjumlah 26 butir soal. Butir soal tersebut adalah butir soal nomor 2, 6, 9, 10, 11, 12, 13, 14, 15, 16, 17, 20, 21, 24, 26, 27, 28, 29, 30, 31, 32, 33, 35, 36, 37, dan 38.

Tabel 3. Butir Soal yang Layak Digunakan

No	Indikator Materi	Nomor Soal	Jenjang Kognitif
1.	Membedakan GLB, GLBB, gerak vertikal, dan GJB	2	C2
2.	Menerapkan persamaan yang terkait dengan GLB untuk menyelesaikan	6, 9 10, 11,	C3 C4

No	Indikator Materi	Nomor Soal	Jenjang Kognitif
	masalah sederhana	12, 14	
		13	C5
3.	Menerapkan persamaan yang terkait dengan GLBB untuk menyelesaikan masalah sederhana	15, 16, 17	C3
		20, 21	C4
		24, 26	C5
4.	Menerapkan persamaan yang terkait dengan gerak vertikal untuk menyelesaikan masalah sederhana	27, 28, 29, 30, 31, 32, 33	C3
		35, 36	C4
		37, 38	C5
5.	Menerapkan persamaan yang terkait dengan GJB untuk menyelesaikan masalah sederhana		C3

SIMPULAN DAN SARAN

Simpulan

Berdasarkan pembahasan yang telah dipaparkan, dapat disimpulkan secara umum bahwa instrumen tes yang dikembangkan dan dianalisis menggunakan Model Rasch dapat digunakan untuk mengukur kemampuan peserta didik. Selain itu, menghasilkan 26 butir soal yang telah memenuhi seluruh karakteristik *item* berdasarkan Model Rasch.

Saran

Agar penelitian menjadi lebih baik, peneliti menyarankan, (a) berkoordinasi dengan guru di sekolah tempat melakukan penelitian bahwa hasil dari instrumen tes yang dikerjakan oleh peserta didik akan dimasukkan ke dalam penilaian atau nilai rapor agar peserta didik lebih serius dalam mengerjakan instrumen tes, (b) menyediakan angket untuk peserta didik yang mengerjakan instrumen tes untuk memperoleh data tambahan mengenai kualitas instrumen tes yang dikembangkan.

DAFTAR RUJUKAN

- Akhtar, H. (2017). *Analisis Item Menggunakan Winstep*. Retrieved from https://www.semestapsikometrika.com/2017/07/analisis-item-menggunakan-winstep_29.html?m=1.
- Akhtar, H. (2017). *Berkenalan dengan Model Rasch*. Retrieved from <https://www.semestapsikometrika.com/2017/07/berkenalan-dengan-rasch-model.html>.
- Alagumalai, S. dkk. (2005). *Applied Rasch Measurement: A Book of Exemplars*. Dordrecht: Springer.
- Arifin, Z. (2009). *Evaluasi Pembelajaran*. Bandung: PT Remaja Rosdakarya.
- Arifin, Z. (2011). *Penelitian Pendidikan Metode dan Paradigma Baru*. Bandung: PT. Remaja Rosdakarya.
- Arikunto, S. 2010. *Dasar-dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Arikunto, S. (2012). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Azwar, S. (2012). *Reabilitas dan Validitas*. Yogyakarta: Pustaka Pelajar.
- Bond, T. G. dan Fox C. M. (2015). *Applying the Rasch Model Fundamental Measurement in the Human Sciences, Third Edition*. New York: Routledge.
- Boone, W. J. dkk. (2014). *Rasch Analysis in the Human Sciences*. London: Springer.
- Buhari, B. (2011). *Four-D (Model Pengembangan Perangkat Pembelajaran dari Thiagarajan, dkk)*. Retrieved from <https://bustangbuhari.wordpress.com/20>

- 11/08/25/four-d-model-model-pengembangan-perangkat-pembelajaran-dari-thiagarajan-dkk/.
- Cavanagh, R. F. dan Waugh, R. F. (2011). *Applications of Rasch Measurement in Learning Environments Research*. Netherlands: Sense Publishers
- Hambleton, R. H and Russell W. J. (1993). An NCME Instructional Module. *Comparison of Classical Theory and Item Response Theory and Their Applications to Test Development*. Volume 12: 38-47.
- Jihad, A dan Haris. (2013). *Evaluasi Pembelajaran*. Yogyakarta: Multi Presindo.
- Mardapi, D. (2012). *Pengukuran Penilaian dan Evaluasi Pendidikan*. Yogyakarta: Nuha Medika.
- Purwanto. (2008). *Evaluasi Hasil Belajar*. Yogyakarta: Pustaka Pelajar.
- Qasem, M. A. N, (2013). *Journal of Research and Method in Education*. A Comparative Study of Classical Theory (Ct) and Item Response Theory (Irt) In Relation To Various Approaches of Evaluating the Validity and Reliability of Research Tools. Volume 3(5): 77-81.
- Sudaryono. (2012). *Dasar-Dasar Evaluasi Pembelajaran*. Yogyakarta: Graha Ilmu.
- Sugiyono. (2014). *Metode Penelitian Pendidikan*. Bandung: Alfabeta.
- Sukardi, H. M. (2012). *Evaluasi Pendidikan Prinsip & Operasionalnya*. Jakarta: Bumi Aksara.
- Sumintono, B. dan Widhiarso, W. (2013). *Aplikasi Model Rasch Untuk Penelitian Ilmu-Ilmu Sosial*. Cimahi: Trim Komunikata.
- Sumintono, B. dan Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch Pada Assessment Pendidikan*. Cimahi: Trim Komunikata.
- Sutrisno, L. (2008). *Remediation of Weaknesses of Physics Concepts*. Pontianak: Untan Press.